Journal of Multivariate Analysis **82**, 445–456 (2002)
doi:10.1006/jmva.2001.2026,

Efficient Estimation of Two Seemingly Unrelated Regression Equations

Aiyi Liu

Georgetown University Medical Center
E-mail: Liua1@gunet.georgetown.edu

Received February 28, 2000; published online December 19, 2001

We derive simpler expressions under a certain structure of design matrices for the two-stage Aitken estimates of the regression coefficients of two seemingly unrelated regression equations. The estimates are shown to have smaller variance than the ordinary least squares estimates for sufficiently large samples.

AMS 1991 subject classifications: 62J05; 62F10.

1. INTRODUCTION

We consider a system of two seemingly unrelated regression (SUR) equations initially introduced by Zellner [6, 7],

$$\mathbf{y}_i = \mathbf{X}_i \beta_i + \varepsilon_i \qquad (i = 1, 2), \tag{1.1}$$

where in the ith equation $\mathbf{y}_i$ is $n \times 1$ vector of observations; $\mathbf{X}_i$ is $n \times p_i$ matrix with rank p_i; β_i is $p_i \times 1$ vector of unknown coefficients; and ε_i is $n \times 1$ vector of disturbances. The rows of $(\varepsilon_1, \varepsilon_2)$ are assumed to be independently distributed, each has a bivariate normal distribution $\mathbf{N}_2(\mathbf{0}, \Sigma)$, where $\Sigma = (\sigma_{ij})$ is a 2×2 unknown positive definite matrix and $\sigma_{12} \neq 0$. We further assume that $n - p_1 - p_2 > 4$ so that valid inference on parameters can be made. We focus on estimating the regression coefficients β_1; β_2 can be estimated accordingly.

Mixing the two equations in (1.1) yields an ordinary linear model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{1.2}$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

0047-259X/01 $35.00

The disturbances ε has mean vector $\mathbf{0}$ and dispersion matrix $\Sigma \otimes \mathbf{I}$, where $\otimes$ denotes the Kronecker product of matrices. The best linear unbiased estimate of β, obtained from the augmented model (1.2), is then

$$\mathbf{b}(\Sigma) = (\mathbf{b}_1'(\Sigma), \mathbf{b}_2'(\Sigma))' = (\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\, \mathbf{X})^{-1}\, \mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\, \mathbf{y}. \tag{1.3}$$

This estimate is, however, not feasible because Σ is not known. Replacing the unknown Σ by its unrestricted estimate (Zellner [7], Revankar [4]) $\frac{1}{n} S$ where

$$S = (s_{ij}) = \mathbf{Y}'(\mathbf{I} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-}\, \tilde{\mathbf{X}}')\, \mathbf{Y}, \tag{1.4}$$

with $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2)$, $\tilde{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2)$, yields the two-stage Aitken estimate of β

$$\mathbf{b}(S) = (\mathbf{b}_1'(S), \mathbf{b}_2'(S))' = (\mathbf{X}'(S^{-1} \otimes \mathbf{I})\, \mathbf{X})^{-1}\, \mathbf{X}'(S^{-1} \otimes \mathbf{I})\, \mathbf{y}. \tag{1.5}$$

Since S is a consistent estimate of Σ, asymptotically the two-stage estimate $\mathbf{b}_1(S)$ in (1.5) is more efficient than the ordinary least squares estimate

$$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\, \mathbf{X}_1'\mathbf{y}_1. \tag{1.6}$$

For small n, however, this may not be true (see (1.10) below), and for how large n that the superiority of $\mathbf{b}_1(S)$ over $\mathbf{b}_1$ continues to hold is not known.

Under the assumption that the column space of $\mathbf{X}_1$ is orthogonal to that of $\mathbf{X}_2$, i.e., $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, Zellner [6, 7] showed that the two-stage Aitken estimate $\mathbf{b}_1(S)$ of β_1 reduces to a much simpler form as

$$\underline{\mathbf{b}_1} = \mathbf{b}_1 - \frac{s_{12}}{s_{22}}\, (\mathbf{X}_1'\mathbf{X}_1)^{-1}\, \mathbf{X}_1'\mathbf{y}_2. \tag{1.7}$$

Revankar [4, 5] showed that if $\mathbf{X}_2$ is a proper subset of $\mathbf{X}_1$, then $\mathbf{b}_1(S)$ is simplified as

$$\hat{\beta}_1(S) = \mathbf{b}_1 - \frac{s_{12}}{s_{22}}\, (\mathbf{X}_1'\mathbf{X}_1)^{-1}\, \mathbf{X}_1'\mathbf{N}_2\mathbf{y}_2, \tag{1.8}$$

where throughout this paper, $\mathbf{P}_i$ and $\mathbf{N}_i$ are defined as

$$\mathbf{P}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\, \mathbf{X}_i', \qquad \mathbf{N}_i = \mathbf{I} - \mathbf{P}_i. \tag{1.9}$$

Note that $\mathbf{P}_i$ and $\mathbf{N}_i$ both are idempotent matrices, orthogonal to each other (Rao [3]).

Revankar [4] further showed that the dispersion matrix of (1.8) is

$$\operatorname{cov}[\hat{\beta}_1(S)] = \sigma_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1} - \sigma_{11}\left[\rho^2 - \frac{1-\rho^2}{n'-2}\right](\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}, \tag{1.10}$$

where $n' = n - p_1 (> 2)$, and $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$, the correlation coefficient between $\mathbf{y}_1$ and $\mathbf{y}_2$.

$\hat{\beta}_1(S)$ and $\mathbf{b}_1$ both are unbiased estimates of β_1, regardless of the two design matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. For a fixed ρ, $\hat{\beta}_1(S)$ is more efficient than $\mathbf{b}_1$ if $n' - 1 > 1/\rho^2$.

The paper is arranged as follows. In Section 2 we present some preliminary results on the unrestricted estimate S. In Section 3 we further simplify the two-stage Aitken estimate $\mathbf{b}_1(S)$ under more general structures. In Section 4 we derive sufficient conditions under which the simplified estimate is more efficient than the ordinary least squares estimate.

2. SOME PRELIMINARY RESULTS

In this section we derive some preliminary properties of the unrestricted estimate S defined in (1.4). These results are used to study the performance of the estimates.

LEMMA 2.1. *Let r be the rank of the matrix $\tilde{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2)$. Define*

$$\tilde{n} = n - r. \tag{2.1}$$

Then S has a Wishart distribution $\mathbf{W}_2(\tilde{n}, \Sigma)$ and s_{ij} is independent of $\mathbf{X}_1'\mathbf{y}_1$ and $\mathbf{X}_1'\mathbf{N}_2\mathbf{y}_2$. Hence s_{ij} is also independent of $\mathbf{P}_1\mathbf{N}_2\mathbf{y}_2$ and $\mathbf{X}_1'\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1$.

Proof. The proof follows immediately since the disturbances are normally distributed. See Rao [3] and Muirhead [1]. ∎

LEMMA 2.2. *For the 2×2 matrix $S = (s_{ij})$ in (1.4)*

$$\frac{s_{12}}{s_{11}} = \sigma_{11}^{-\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}\,t_1\right), \tag{2.2}$$

$$\frac{s_{12}}{s_{22}} = \sigma_{11}^{\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}\,t_2\right), \tag{2.3}$$

where t_i $(i=1,2)$ has a Student's t-distribution with $\tilde{n}$ degrees of freedom (see Rao [3]).

Proof. We first show (2.2). Define

$$\mathbf{C}=\frac{1}{\sqrt{1-\rho^2}}\begin{pmatrix}\sigma_{11}^{-\frac{1}{2}}\sqrt{1-\rho^2} & -\rho\sigma_{11}^{-\frac{1}{2}}\\ 0 & \sigma_{22}^{-\frac{1}{2}}\end{pmatrix}. \tag{2.4}$$

Then

$$\mathbf{C}'\Sigma\mathbf{C}=\mathbf{I}. \tag{2.5}$$

Further let

$$\mathbf{A}=\mathbf{C}'S\mathbf{C}. \tag{2.6}$$

Then $\mathbf{A}$ has a Wishart distribution $\mathbf{W}_2(\tilde{n},\mathbf{I})$. Following the well known Bartlett's decomposition (see Muirhead [1]), put $\mathbf{A}=\mathbf{T}'\mathbf{T}$, where

$$\mathbf{T}=\begin{pmatrix}t_{11} & t_{12}\\ 0 & t_{22}\end{pmatrix}. \tag{2.7}$$

Then the t_{ij}'s are independent and $t_{12}\sim \mathbf{N}(0,1)$, $t_{11}\sim\chi^2_{\tilde{n}}$, and $t_{22}\sim\chi^2_{\tilde{n}-1}$. It follows from the equation $\mathbf{T}'\mathbf{T}=\mathbf{C}'S\mathbf{C}$ that

$$t_{11}^2=s_{11}\sigma_{11}^{-1}, \tag{2.8}$$

$$t_{12}t_{11}=\frac{1}{\sqrt{1-\rho^2}}\left(s_{12}\sigma_{11}^{-\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}-s_{11}\rho\sigma_{11}^{-1}\right). \tag{2.9}$$

Hence,

$$\frac{t_{12}}{t_{11}}=\frac{1}{\sqrt{1-\rho^2}}\left(\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}\frac{s_{12}}{s_{11}}-\rho\right), \tag{2.10}$$

that is,

$$\frac{s_{12}}{s_{11}}=\sigma_{11}^{-\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}\left(\rho+\frac{t_{12}}{t_{11}/\tilde{n}}\,\frac{\sqrt{1-\rho^2}}{\tilde{n}}\right). \tag{2.11}$$

Let $t_1=t_{12}/(t_{11}/\tilde{n})$. Then t_1 has a Student's t-distribution with $\tilde{n}$ degrees of freedom. This proves (2.2).

Proof of (2.3) follows similarly by substituting $\mathbf{D}$ for $\mathbf{C}$, where

$$\mathbf{D}=\frac{1}{\sqrt{1-\rho^2}}\begin{pmatrix} 0 & \sigma_{11}^{-\frac{1}{2}} \\ \sigma_{22}^{-\frac{1}{2}}\sqrt{1-\rho^2} & -\rho\sigma_{22}^{-\frac{1}{2}} \end{pmatrix}. \tag{2.12}$$

Moments of a Students's t-distribution are well known (see Patel *et al.* [2]). They are presented in Lemma 2.3. ∎

LEMMA 2.3. *Let $t_{\tilde{n}}$ be a Student's t-distribution with $\tilde{n}$ degrees of freedom. Then for any $k<\tilde{n}$,*

$$\mathbf{E}[t_{\tilde{n}}^k]=\begin{cases} \tilde{n}^{2/k}\dfrac{1\cdot 3\cdot 5\cdots(k-1)}{(\tilde{n}-2)(\tilde{n}-4)\cdots(\tilde{n}-k)}, & k \text{ being even} \\ 0, & k \text{ being odd.} \end{cases}$$

Using Lemmas 2.2 and 2.3 one can easily obtain any moments of s_{12}/s_{ii} $(i=1,2)$. We list some moments used later in the present paper:

$$\mathbf{E}\left[\frac{s_{12}}{s_{ii}}\right]=\frac{\sigma_{12}}{\sigma_{ii}} \qquad (i=1,2), \tag{2.13}$$

$$\mathbf{E}\left[\left(\frac{s_{12}}{s_{22}}\right)^2\right]=\frac{\sigma_{11}}{\sigma_{22}}\rho^2+\frac{\sigma_{11}}{\sigma_{22}}(1-\rho^2)\frac{1}{n-r-2}. \tag{2.14}$$

3. SIMPLER VERSION OF TWO-STAGE ESTIMATES

We now derive a simpler version of two-stage estimates (1.5) under a very general structure. This simpler version allows us to study the performance of such estimates in a more detailed way in Section 4.

THEOREM 3.1. *Let the two-stage estimate $\mathbf{b}_1(S)$ be defined as in (1.5), and $\mathbf{P}_i$ and $\mathbf{N}_i$ $(i=1,2)$ as in (1.9). Define*

$$\tilde{\beta}_1(S)=\mathbf{b}_1-\frac{s_{12}}{s_{22}}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{y}_2+\hat{\rho}^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1, \tag{3.1}$$

where $\mathbf{b}_1$ is the ordinary least squares estimate of β_1, $\hat{\rho}=s_{12}/(s_{11}s_{22})^{1/2}$ is the sample correlation coefficient between the two observable variables $\mathbf{y}_1$ and $\mathbf{y}_2$. Then $\tilde{\beta}_1(S)=\mathbf{b}_1(S)$ if and only if

$$\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2=\mathbf{0}. \tag{3.2}$$

Proof. The two-stage Aitken estimates $\mathbf{b}_1(S)$ and $\mathbf{b}_2(S)$ in (1.5) satisfy the equation

$$\mathbf{X}'(S^{-1}\otimes\mathbf{I})\,\mathbf{X}\begin{pmatrix}\mathbf{b}_1(S)\\ \mathbf{b}_2(S)\end{pmatrix}=\mathbf{X}'(S^{-1}\otimes\mathbf{I})\,\mathbf{y}, \tag{3.3}$$

that is,

$$\begin{pmatrix} s_{22}\mathbf{X}_1'\mathbf{X}_1 & -s_{21}\mathbf{X}_1'\mathbf{X}_2\\ -s_{12}\mathbf{X}_2'\mathbf{X}_1 & s_{11}\mathbf{X}_2'\mathbf{X}_2\end{pmatrix}\begin{pmatrix}\mathbf{b}_1(S)\\ \mathbf{b}_2(S)\end{pmatrix}=\begin{pmatrix} s_{22}\mathbf{X}_1' & -s_{21}\mathbf{X}_1'\\ -s_{12}\mathbf{X}_2' & s_{11}\mathbf{X}_2'\end{pmatrix}\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}.$$

Left-multiplying both sides of the above equation by the matrix

$$\begin{pmatrix}\mathbf{I} & \frac{s_{12}}{s_{11}}\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}$$

yields

$$\begin{aligned}\left(s_{22}\mathbf{X}_1'\mathbf{X}_1-\frac{s_{12}^2}{s_{11}}\mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1\right)\mathbf{b}_1(S)&=s_{22}\mathbf{X}_1'\mathbf{y}_1-s_{12}\mathbf{X}_1'\mathbf{N}_2\mathbf{y}_2\\ &\quad+\frac{s_{12}^2}{s_{11}}\mathbf{X}_1'\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1-\frac{s_{12}^2}{s_{11}}\mathbf{X}_1'\mathbf{y}_1+\frac{s_{12}^2}{s_{11}}\mathbf{X}_1'\mathbf{N}_2\mathbf{P}_1\mathbf{y}_1.\end{aligned} \tag{3.4}$$

Note that $\tilde{\beta}_1(S)=\mathbf{b}_1(S)$ if and only if $\tilde{\beta}_1(S)$ is a solution for $\mathbf{b}_1(S)$ to Eq. (3.4). Replacing $\mathbf{b}_1(S)$ in (3.4) by the right side of Eq. (3.1) we obtain

$$\begin{aligned}&\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{y}_1-\frac{s_{12}}{s_{22}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{y}_2+\frac{s_{12}^2}{s_{11}s_{22}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1\\ &\quad=\mathbf{X}_1'\mathbf{y}_1-\mathbf{X}_1'\mathbf{N}_2\mathbf{P}_1\mathbf{y}_1.\end{aligned} \tag{3.5}$$

Since $\mathbf{P}_2+\mathbf{N}_2=\mathbf{I}$, and $\mathbf{X}_1'\mathbf{P}_1=\mathbf{X}_1'$, (3.5) becomes

$$\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{y}_2=\frac{s_{12}}{s_{11}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1,$$

that is,

$$\left(\frac{s_{12}}{s_{11}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1,\ -\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\right)\mathbf{y}=\mathbf{0}. \tag{3.6}$$

Hence $\tilde{\beta}_1(S) = \mathbf{b}_1(S)$ if and only if (3.6) holds. It remains to show that (3.2) is a necessary and sufficient condition for (3.6).

Sufficiency. Suppose (3.2) holds, i.e., $\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2 = \mathbf{0}$. Then $\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2 = \mathbf{0}$, and hence Eq. (3.6) holds.

Necessity. Suppose (3.6) holds. Define

$$\delta = \left(\frac{s_{12}}{s_{11}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1,\ -\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\right)\mathbf{y},$$

the left side of Eq. (3.6). Then $\mathbf{E}[\delta] = 0$, and $\operatorname{cov}[\delta] = 0$. Note that s_{ij} is independent of both $\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1\mathbf{y}_1$ and $\mathbf{P}_1\mathbf{N}_2\mathbf{y}_2$. It follows that the dispersion matrix of δ is

$$\begin{aligned}
\mathbf{0} &= \operatorname{cov}[\delta] = \mathbf{E}[\delta\delta'] = \mathbf{E}\{\mathbf{E}[\delta\delta' \mid s_{ij}]\}\\
&= \left(\mathbf{E}\left[\frac{s_{12}}{s_{11}}\right]\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1,\ -\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\right)(\Sigma^{-1}\otimes\mathbf{I})\\
&\quad\times\begin{pmatrix}\mathbf{E}\left[\dfrac{s_{12}}{s_{11}}\right]\mathbf{N}_1\mathbf{N}_2\mathbf{P}_1\mathbf{P}_2\mathbf{X}_1\\ -\mathbf{N}_2\mathbf{P}_1\mathbf{P}_2\mathbf{X}_1\end{pmatrix}\\
&= \left(\frac{\sigma_{12}}{\sigma_{11}}\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\mathbf{N}_1,\ -\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2\right)(\Sigma^{-1}\otimes\mathbf{I})\begin{pmatrix}\dfrac{\sigma_{12}}{\sigma_{11}}\mathbf{N}_1\mathbf{N}_2\mathbf{P}_1\mathbf{P}_2\mathbf{X}_1\\ -\mathbf{N}_2\mathbf{P}_1\mathbf{P}_2\mathbf{X}_1\end{pmatrix}.
\end{aligned}$$

It follows that $\mathbf{X}_1'\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2 = \mathbf{0}$, since Σ is a positive definite matrix. This completes the proof.

If $\mathbf{X}_2$ is a subset of $\mathbf{X}_1$, then $\mathbf{P}_2\mathbf{P}_1 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{P}_1 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2' = \mathbf{P}_2$, implying $\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{N}_2 = \mathbf{0}$. If $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, then (3.2) obviously holds. Hence the structures of the two design matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, considered by Zellner [6, 7] and Revankar [4], are special cases of (3.2).

4. FINITE SAMPLE EFFICIENCY OF THE TWO-STAGE ESTIMATES

In this section we study the relative finite sample efficiency of the two-stage estimate (3.1), as compared to the ordinary least squares estimate $\mathbf{b}_1$. Since all the estimates of interest are unbiased estimates, we need only compare their dispersion matrices.

Using Lemma 2.1 of Section 2 and by straightforward calculation we obtain

$$\begin{aligned}
\operatorname{cov}[\tilde{\beta}_1(S)] &= \mathbf{E}[(\tilde{\beta}_1(S)-\beta_1)(\tilde{\beta}_1(S)-\beta_1)'] \\
&= \mathbf{E}\{\mathbf{E}[(\tilde{\beta}_1(S)-\beta_1)(\tilde{\beta}_1(S)-\beta_1)' \mid S]\} \\
&= \operatorname{cov}[\mathbf{b}_1]-\sigma_{11}\left(\rho^2-\frac{1-\rho^2}{n'-2}\right)(\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\
&\quad -\sigma_{11}\left(2\frac{\sigma_{12}}{\sigma_{11}}\mathbf{E}\left[\frac{s_{12}^3}{s_{11}s_{22}^2}\right]-E\left[\frac{s_{12}^4}{s_{11}^2s_{22}^2}\right]\right) \\
&\quad \times(\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{N}_2\mathbf{N}_1\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \\
&= \operatorname{cov}[\mathbf{b}_1]-\sigma_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{N}_2\,\Delta\mathbf{N}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1},
\end{aligned}$$

where

$$\Delta=\left(\rho^2-\frac{1-\rho^2}{n'-2}\right)\mathbf{I}+\lambda\mathbf{N}_1, \tag{4.1}$$

with

$$\lambda=2\frac{\sigma_{12}}{\sigma_{11}}\mathbf{E}\left[\frac{s_{12}^3}{s_{11}s_{22}^2}\right]-\mathbf{E}\left[\frac{s_{12}^4}{s_{11}^2s_{22}^2}\right]. \tag{4.2}$$

Note that Δ is a symmetric matrix with eigenvalues either $\rho^2-(1-\rho^2)/(n'-2)$ or $\lambda+\rho^2-(1-\rho^2)/(n'-2)$. We therefore conclude that under the following two conditions,

$$n\geqslant p_1+\frac{1}{\rho^2}+1, \tag{4.3}$$

and

$$\lambda=2\frac{\sigma_{12}}{\sigma_{11}}\mathbf{E}\left[\frac{s_{12}^3}{s_{11}s_{22}^2}\right]-\mathbf{E}\left[\frac{s_{12}^4}{s_{11}^2s_{22}^2}\right]\geqslant 0, \tag{4.4}$$

the two-stage estimate $\tilde{\beta}_1(S)$ is at least as efficient as the ordinary least squares estimate $\mathbf{b}_1$. This is so since under these two conditions, Δ is non-negative definite and hence $\operatorname{cov}[\tilde{\beta}_1(S)]-\operatorname{cov}[\mathbf{b}_1]$ is non-postitive definite.

It is worth noting that (4.3) is a sufficient condition that Revankar's [4] estimate is more efficient than the ordinary least squares estimate.

Condition (4.4) is practically difficult to verify. We now derive some simpler sufficient conditions under which $\tilde{\beta}_1(S)$ is superior to $\mathbf{b}_1$.

THEOREM 4.1. *The estimate $\tilde{\beta}_1(S)$ is at least as efficient as $\mathbf{b}_1$ if*

$$n \geqslant p_1 + p_2 + \frac{3}{\rho^{\frac{8}{3}}} + 4. \tag{4.5}$$

Proof. Obviously (4.5) implies (4.3) since $0 \leqslant \rho \leqslant 1$. We now show that (4.5) implies $\lambda > 0$. Let r and $\tilde{n}$ be defined as in Lemma 2.1. It follows from (2.2) and (2.3) that

$$\begin{aligned}\frac{s_{12}^3}{s_{11}s_{22}^2} &= \frac{s_{12}}{s_{11}}\left(\frac{s_{12}}{s_{22}}\right)^2\\ &= \sigma_{11}^{-\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_1\right)\left[\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_2\right)\right]^2\\ &= \sigma_{11}^{\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}\left[\rho^3 + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}\rho^2 t_1 + 2\rho^2\frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_2\right.\\ &\quad \left. + 2\rho\frac{1-\rho^2}{\tilde{n}}t_1t_2 + \rho\frac{1-\rho^2}{\tilde{n}}t_2^2 + \left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}t_1t_2^2\right],\end{aligned}$$

and

$$\begin{aligned}\frac{s_{12}^4}{s_{11}^2s_{22}^2} &= \left(\frac{s_{12}}{s_{11}}\right)^2\left(\frac{s_{12}}{s_{22}}\right)^2\\ &= \left[\sigma_{11}^{-\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_1\right)\right]^2\left[\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{-\frac{1}{2}}\left(\rho + \frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_2\right)\right]^2\\ &= \rho^4 + 2\rho^3\frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_1 + \rho^2\frac{1-\rho^2}{\tilde{n}}t_1^2\\ &\quad + 2\rho^3\frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}t_2 + 4\rho^2\frac{1-\rho^2}{\tilde{n}}t_1t_2 + 2\rho\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}t_1^2t_2\\ &\quad + \rho^2\frac{1-\rho^2}{\tilde{n}}t_2^2 + 2\rho\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}t_1t_2^2 + \left(\frac{1-\rho^2}{\tilde{n}}\right)^2t_1^2t_2^2.\end{aligned}$$

Note that t_1 and t_2 have the same distribution as $t_{\tilde{n}}$ and $\mathbf{E}[t_{\tilde{n}}]=0$, we have

$$\begin{aligned}
\lambda &= 2\frac{\sigma_{12}}{\sigma_{11}}\mathbf{E}\left[\frac{s_{12}^3}{s_{11}s_{22}^2}\right]-\mathbf{E}\left[\frac{s_{12}^4}{s_{11}^2s_{22}^2}\right]\\
&=2\rho\left(\rho^3+3\rho^2\frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}\mathbf{E}[t_{\tilde{n}}]+\rho\frac{1-\rho^2}{\tilde{n}}\mathbf{E}[t_{\tilde{n}}^2]\right.\\
&\quad\left.+2\rho\frac{1-\rho^2}{\tilde{n}}\mathbf{E}[t_1t_2]+\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}\mathbf{E}[t_1t_2^2]\right)\\
&\quad-\left(\rho^4+4\rho^3\frac{\sqrt{1-\rho^2}}{\sqrt{\tilde{n}}}\mathbf{E}[t_{\tilde{n}}]+2\rho^2\frac{1-\rho^2}{\tilde{n}}\mathbf{E}[t_{\tilde{n}}^2]+4\rho^2\frac{1-\rho^2}{\tilde{n}}\mathbf{E}[t_1t_2]\right.\\
&\quad\left.+2\rho\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}\mathbf{E}[t_1^2t_2+t_1t_2^2]+\left(\frac{1-\rho^2}{\tilde{n}}\right)^{2}\mathbf{E}[t_1^2t_2^2]\right)\\
&=\rho^4-\left(\frac{1-\rho^2}{\tilde{n}}\right)^{2}\mathbf{E}[t_1^2t_2^2]-2\rho\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}\mathbf{E}[t_1^2t_2].
\end{aligned}\tag{4.6}$$

Use the Cauchy–Schwarz inequality and Lemma 3.3 to obtain

$$\mathbf{E}[t_1^2t_2^2)]\leqslant(\mathbf{E}[t_1^4]\,\mathbf{E}[t_2^4])^{\frac{1}{2}}=\mathbf{E}[t_{\tilde{n}}^4]=\frac{3\tilde{n}^2}{(\tilde{n}-2)(\tilde{n}-4)},$$

and

$$\mathbf{E}[|t_1^2t_2|]\leqslant(\mathbf{E}[t_1^4]\,\mathbf{E}[t_2^2])^{\frac{1}{2}}=(\mathbf{E}[t_{\tilde{n}}^4]\,\mathbf{E}[t_{\tilde{n}}^2])^{\frac{1}{2}}=\frac{\sqrt{3}\,\tilde{n}^{3/2}}{(\tilde{n}-2)\sqrt{\tilde{n}-4}}.$$

Replacing the two expectation terms in (4.6) by their upper bounds derived above yields

$$\begin{aligned}
\lambda &\geqslant \rho^4-\left(\frac{1-\rho^2}{\tilde{n}}\right)^{2}\frac{3\tilde{n}^2}{(\tilde{n}-2)(\tilde{n}-4)}-2\,|\rho|\left(\frac{1-\rho^2}{\tilde{n}}\right)^{\frac{3}{2}}\frac{\sqrt{3}\,\tilde{n}^{3/2}}{(\tilde{n}-2)\sqrt{\tilde{n}-4}}\\
&>\rho^4-\frac{3}{(\tilde{n}-2)(\tilde{n}-4)}-\frac{\sqrt{3}}{(\tilde{n}-2)\sqrt{\tilde{n}-4}}>\rho^4-\frac{3\sqrt{3}}{(\tilde{n}-4)^{\frac{3}{2}}}.
\end{aligned}\tag{4.7}$$

The second inequality in (4.7) holds since $0\leqslant 1-\rho^2\leqslant 1$, and $2\,|\rho|(1-\rho^2)^{1/2}=2\sqrt{\rho^2(1-\rho^2)}\leqslant 1$; the last inequality holds since $\tilde{n}-2>\tilde{n}-4>\sqrt{\tilde{n}-4}$, and $3+\sqrt{3}<3\sqrt{3}$.

TABLE I

Values of n_0 for Selected ρ and p

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n_0	1397	224	79	39	24	16	12	10	8

Now suppose (4.5) holds. Since $r = rank(\tilde{X}) \leqslant p_1 + p_2$, it follows that

$$\tilde{n} = n - r \geqslant n - (p_1 + p_2) \geqslant \frac{3}{\rho^{\frac{8}{3}}} + 4, \tag{4.8}$$

that is,

$$\frac{3\sqrt{3}}{(\tilde{n}-4)^{\frac{3}{2}}} \leqslant \rho^4.$$

Hence $\lambda > 0$. This completes the proof.

Let $n_0 = 3/\rho^{8/3} + 4$. The corresponding values of n_0 based on (4.5) for selected values of ρ are shown in Table I.

It is noted that small ρ corresponds to large n. The range of ρ over which $\tilde{\beta}_1(S)$ performs better than $\mathbf{b}_1$ narrows down as the sample size n gets smaller. If n is fixed, we may decide roughly the range of ρ over which $\tilde{\beta}_1(S)$ performs better. On the other hand, for a given ρ, the estimate $\tilde{\beta}_1(S)$ becomes more efficient than $\mathbf{b}_1$ as n increases. In practical situation when the sample size n is moderate, the choice between $\tilde{\beta}_1(S)$ and $\mathbf{b}_1$ may, in general, base on a prior test for ρ.

ACKNOWLEDGMENTS

The author thanks Professor Edmund Gehan for his continuous support in statistical research. The author also thanks the editor and an anonymous referee for their valuable comments and suggestions that led to an improved paper.

REFERENCES

1. R. J. Muirhead, "Aspects of Multivariate Statistical Theory," Wiley, New York, 1982.
2. J. K. Patel *et al.*, "Handbook of Statistical Distributions," Dekker, New York, 1976.
3. C. R. Rao, "Linear Statistical Inference and Its Applications," Wiley, New York, 1973.
4. N. S. Revankar, Some finite sample results in the context of two seemingly unrelated regression equations, *J. Amer. Statist. Assoc.* **69** (1974), 187–190.

5. N. S. Revankar, Use of restricted residuals in SUR systems: Some finite sample results, *J. Amer. Statist. Assoc.* **71** (1976), 183–188.
6. A. Zellner, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *J. Amer. Statist. Assoc.* **57** (1962), 348–368.
7. A. Zellner, Estimates for seemingly unrelated regression equations: Some exact finite sample results, *J. Amer. Statist. Assoc.* **58** (1963), 977–992.